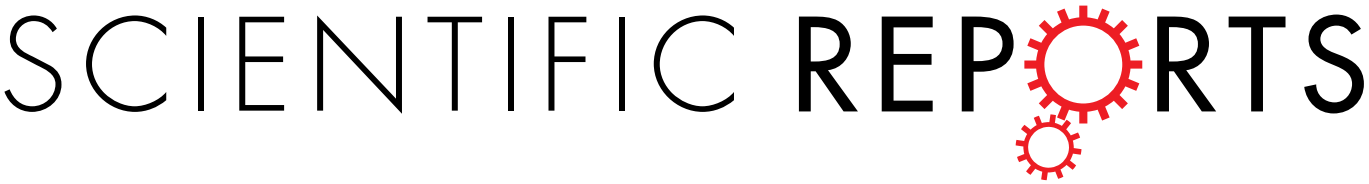
Two universal physical principles shape the topological statistics of real-world networks

Lorimer, Tom ; Gomez, Florian ; Stoop, Ruedi

Abstract: The study of complex networks has pursued an understanding of macroscopic behaviour by focusing on power-laws in microscopic observables. Here, we uncover two universal fundamental physical principles that are at the basis of complex network generation. These principles together predict the generic emergence of deviations from ideal power laws, which were previously discussed away by reference to the thermodynamic limit. Our approach proposes a paradigm shift in the physics of complex networks, toward the use of power-law deviations to infer meso-scale structure from macroscopic observations.

DOI: https://doi.org/10.1038/srep12353



OPEN

Two universal physical principles shape the power-law statistics of real-world networks

Received: 23 April 2015
Accepted: 26 June 2015
Published: 23 July 2015

Tom Lorimer, Florian Gomez & Ruedi Stoop

The study of complex networks has pursued an understanding of macroscopic behaviour by focusing on power-laws in microscopic observables. Here, we uncover two universal fundamental physical principles that are at the basis of complex network generation. These principles together predict the generic emergence of deviations from ideal power laws, which were previously discussed away by reference to the thermodynamic limit. Our approach proposes a paradigm shift in the physics of complex networks, toward the use of power-law deviations to infer meso-scale structure from macroscopic observations.

A recent seminal discovery elucidated that in nature a simple physical principle often rules the growth of 'random networks'. The so called preferential attachment ('the rich get richer') rule leads to complex networks that have properties contrasting those predicted from classical random network theory[1–4]. A fundamental universality principle of physics must be held responsible for this change of paradigm. The preferential attachment principle expresses in our interpretation that for the formation of ensembles, attractive forces that are generally valid over decades of spatial extensions are required (that in physics may involve, e.g., mass, charge). It is this principle that generates the celebrated power laws observed in the distribution of mesoscopic network indicators, such as network degree, connectivity weight[5–8], or neuronal avalanche size[9–11]. A *second* fundamental universality principle of physics that is active at the same time, has, however, passed unnoticed so far. It is the fact that real-world connectivity requires space, and that this space is limited. The question that we address in our work is what the traces of this principle will be, during network formation and regarding the final network. This question has not been answered so far.

Generic network building algorithm

To study this question, we consider a novel generic network building algorithm (our 'primary model') that implements both principles at the most basic level as follows. We start from a connected network of N_0 nodes. With probability p, an 'outside' node, from a finite set of available nodes, is added; alternatively, with probability $1-p$, an attempt is made to construct an 'inside' edge (see below). If an outside node is added, the new node connects to the network by m edges, where the target nodes are sampled according to their degree k (i.e. $\propto k$), following preferential attachment. For an inside edge, two nodes are independently chosen using preferential attachment. If the two chosen nodes are not identical and not already connected, an edge is established, which expresses the second fundamental principle in terms of an 'edge saturation' (at a level defined by p and m, implemented right from the start of the network's growth). The process stops if the set of available nodes is depleted. The algorithm generates undirected topological networks of arbitrary size, void of loops and multiple-edges; examples will be discussed later. Fig. 1 shows the stereotypical degree distribution obtained in this way, exhibiting an extended power-law part of the distribution terminated by a hump (that, upon the network's growth, moves towards larger

Institute of Neuroinformatics and Institute of Computational Science, University of Zurich and ETH Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland. Correspondence and requests for materials should be addressed to R.S. (email: ruedi@ini.phys.ethz.ch)

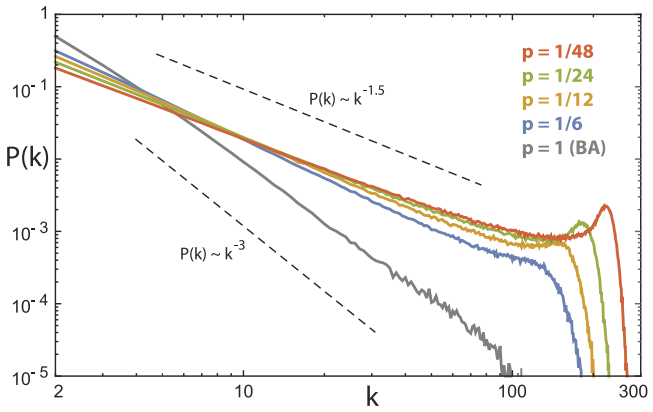


Figure 1. Characteristic degree distributions from the two key principles (for different values of parameter p and fixed parameter $m=2$; the effect of m is exhibited in Figs 3 and 4). Network size $t=10^3$ nodes, mean of 10^3 realizations. Dashed lines: power-law visual guides. The effect is most saliently expressed for exponents <2, occurring often in gene or protein networks.

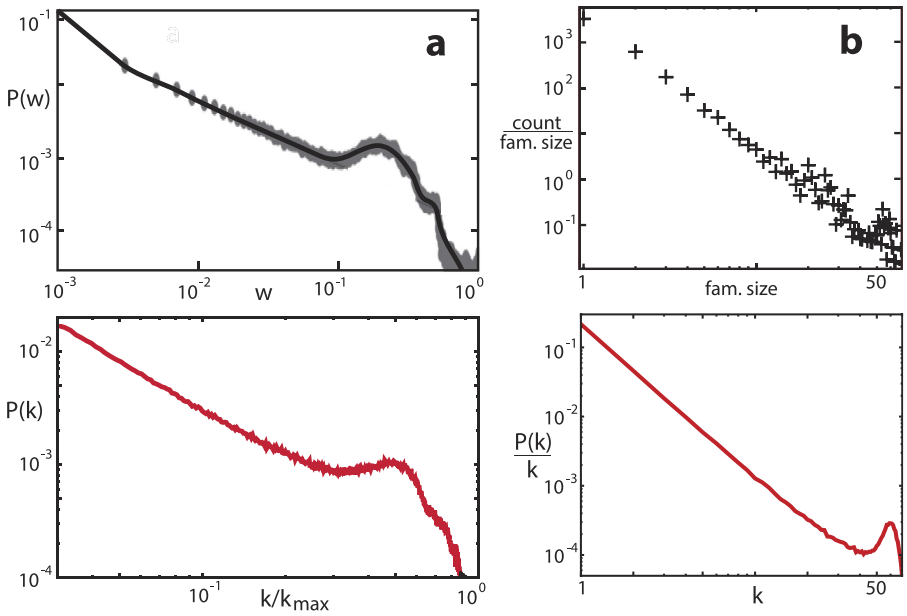


Figure 2. Typical weight and degree distributions, respectively, from experiments, and their qualitative modelling (black: experimental, red: simulation data). (**a**) Network of synchronizing linear phase oscillators (network weight distribution during synchronization)[8]. (**b**) Gene family for *S. cerevisiae*[35] (family size distribution). For the modelling, different (p, m)-models were superimposed for (**a**).

degrees, until the process is stopped by node depletion. The details of how this happens in time are outlined in our 'Statistical modelling' section).

Network properties

While we observe a wide-spread activity to find power-law distributions in all areas of physics, we emphasize that based on the fundamental ingredients necessary in the network building process, only in rare cases will neat power laws be found. Examples of experimental data with the deviations that our key principles predict are shown in Fig. 2. While our real-world examples are often related to biology (mostly because of the great availability of the underlying data, and because of the greater simplicity of the examples), all of our arguments are immediately transferable to physical situations where previous analysis has generally stopped at the preferential attachment level. Our analysis now provides guidelines for inferring from macroscopic measurements the microscopic properties that dominate network growth (cf. Fig. 3, where the 'humpiness' of the distribution $P(k)$ was evaluated as the deviation from the power law $p(k)$ excluding the hump, as $(P(k)-p(k))/p(k)$). This provides an important input for the modelling

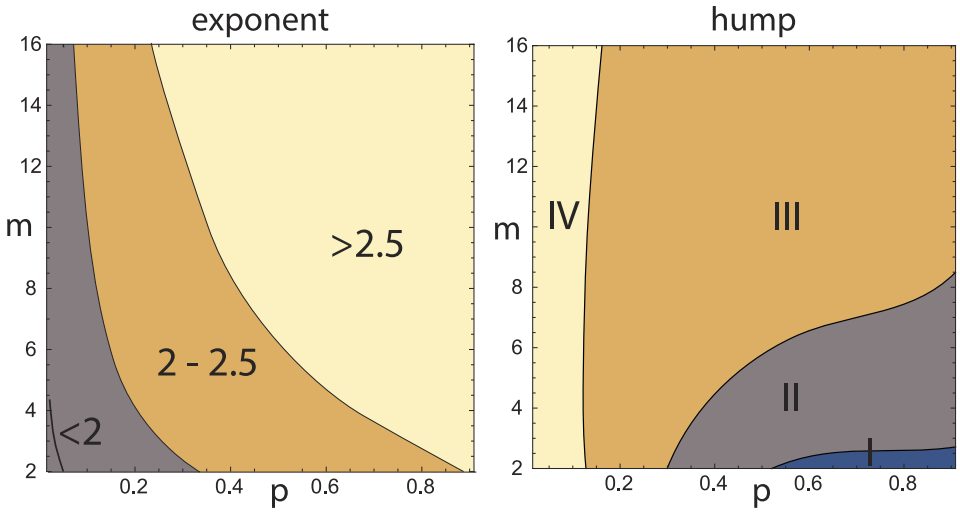


Figure 3. Modelling guidelines: Phase diagram of the humped power law's exponent and 'humpiness' on local parameters (p, m) (see text). Domains of humpiness: I) not resolvable, II minor, III significant, IV salient. Guided by the power-law paradigm, investigations have mostly focused on examples from domains I and II. Network sizes: $t = 10^3$.

of real world systems (see, e.g., the *Drosophila* network example discussed below). By superposition of prototypes with different p and m parameters, more general hump structures can be generated (Fig. 2). This mechanism provides an as yet unexplored link between the macro- and meso-scales that can be invaluable for both the modelling and the further analysis of real-world systems.

In contrast to preferential attachment networks (cf.[12]), a network generated according to the two fundamental physical principles embodied in our primary model, will not necessarily be sparse (this would imply a power-law exponent >2, cf. Fig. 1). Moreover, Dorogovtsev and Mendes' modified preferential attachment algorithm with its double regimes of power-law behaviour[7] also deviates from the fundamental principles that we have worked out. Their model uses a second internal linking process that is always successful in making new connections. In our case it is exactly the edge connection failures (by edge saturation) that define the network structure. Whereas the rate of internal linking in their algorithm accelerates with the network size, our approach does not share this property. The network structures that we obtain depend primarily on parameter p; the obtained distributions are generally unaffected by the network's initial condition (in contrast to Refs 13–15), as long as the initial network size N_0 is sufficiently smaller than the final network size. Previous authors have also studied network shaping by edge depletion[16]. Their algorithm can also produce scale free networks with exponent <2, but excludes saturation, and thus does not show the characteristic hump termination discussed here.

The modelling of biological networks containing a small number of nodes only, is a particular challenge. The example of *Drosophila's* courtship network, a network that is built on observable irreducible acts of body language[17,18] (cf. Figs 4 and 5) illustrates that our approach also successfully masters this challenge (a further discussion of this example is given towards the end of the paper).

Statistical modelling

To better understand how the statistical properties and in particular, saturation, emerge from the model, we focus on a semi-analytical growth description, in which the natural time step t is the addition of one node to the network. The degree distribution from a network growth algorithm is usually determined from a differential equation that describes the rate of addition of new edges to a given node, as a function of the time s at which the node has joined the network[19], i.e. $\frac{\partial k(s,t)}{\partial t} = f(k, s, t)$. For our algorithm, the topological constraint on the addition of inside edges implies that $\frac{\partial k(s,t)}{\partial t}$ can not be determined analytically from the single node information $f(k, s, t)$, but requires the full pairwise connection information of the network encoded in the adjacency matrix at time t, A_t, i.e. $\frac{\partial k(s,t)}{\partial t} = f(k, s, t, A_t)$. To work around this complication, we make the following ansatz. We suppose that the probability of failure while trying to add an inside edge (i, j) to an already chosen node i, can be expressed by a mean field 'saturation' function $F(k, t)$ in terms of the degree k of node i. Furthermore, suppose that the total number of edges present in the network at time t can be approximated by $K(t)$. $F(k, t)$ is then defined as the average probability of a node with degree k, to be already connected to a second node j chosen with $P \propto k_j$. Thus,

$$F(k, t) := \langle F_i(t) \rangle_{k_i = k}, \tag{1}$$

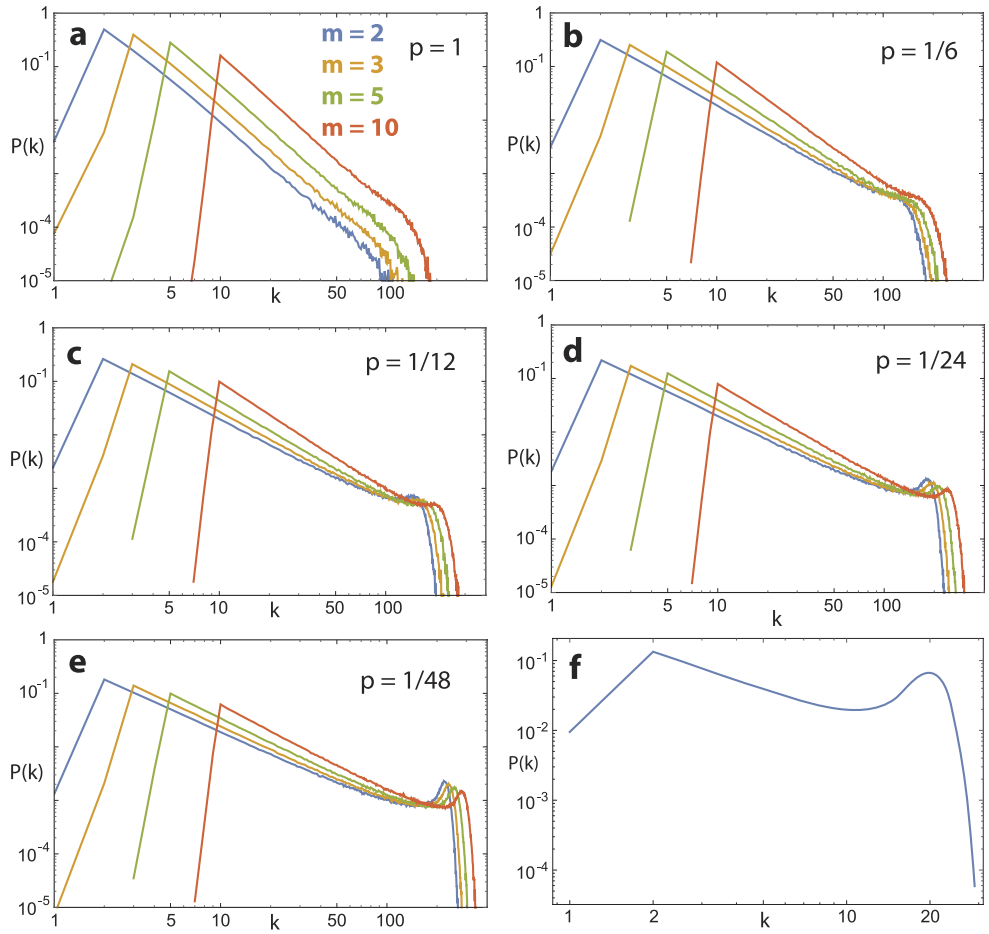


Figure 4. (**a**–**e**) Effect of choice of m on network degree distribution, for different values of p (network size $t = 10^3$ nodes, mean of 10^3 realizations). Increasing m for $p << 1$ increases the influence of the first term in Eq. (3), which increases the exponent by pushing the primary model towards the preferential attachment model. (**f**) Real-world example: Drosophila courtship network's degree distribution (corresponding to the full line in Fig. 5). Degrees $k < m$ have small probability.

where $F_i(t)$ is the probability that node i with degree k_i, is already connected to node j. $F_i(t)$ has then the form

$$F_i(t) := \frac{k_i(t) + \sum_{(i,j)\in E(t)} k_j(t)}{\sum_j k_j(t)}, \tag{2}$$

where $k_i(t)$ accounts for the case where node i would be chosen twice, and the second term is the degree-weighted sum over the nodes to which node i is already connected ($E(t)$ denotes the network's set of edges).

Using this approximation, we can express our algorithm by the rate of addition of new edges to a node of degree $k(s, t)$ as

$$\frac{\partial k(s,t)}{\partial t} = \frac{mk(s,t)}{2K(t)} + \frac{1-p}{p}\frac{k(s,t)}{K(t)}[1 - F(k,t)]. \tag{3}$$

In this case, the network grows out from a connected network of N_0 nodes, with $k(s, s) \approx m$ as the initial condition. The first term on the right hand side of Eq. (3) describes the increase in k due to connection to outside nodes, and the second term describes the addition of inside edges. The whole equation has been rescaled by $\frac{1}{p}$ (cancelling the p in the first term's numerator) such that t corresponds to the number of nodes in the network. As can be easily seen from Eq. (3), our growth algorithm provides two well-known limiting cases. For $p = 1$ we retrieve the preferential attachment growth process[4]. For $p = 0$, the network will not add nodes and must asymptotically become a clique of size N_0. In between, for

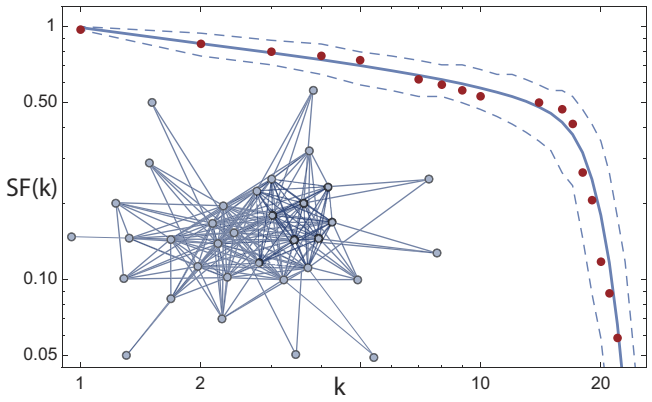


Figure 5. Drosophila courtship language network degree distribution. Survival function $SF(k) := 1 - CDF(k)$, where CDF is the cumulative distribution function (red dots: original data). Solid line: means, dashed lines: 0.05 quantiles, from 1000 realizations of our network growth algorithm ($N = 34, p = \frac{1}{27}$, $m = 2$). Inset: mapped-out *Drosophila* language network.

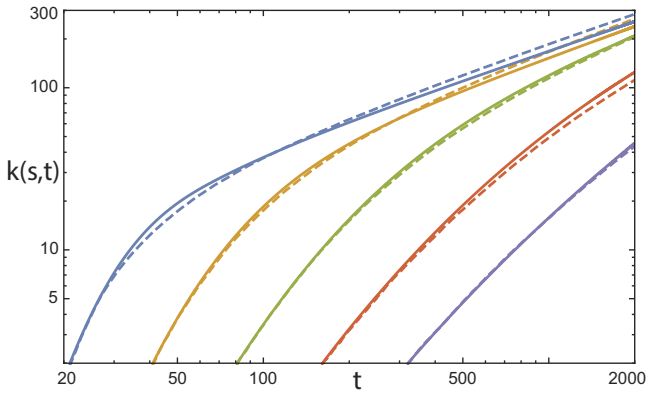


Figure 6. Comparison: Primary model/semi-analytical description. Degree evolution $k(s, t)$ of nodes entering the network at $s = 21, 41, 81, 161, 321$. Mean of 10^3 primary model realizations (dashed), compared with numerical integration of Eq. (3) (solid).

$p << 1$, the second term dominates, which renders the network more dense, and produces the large deviation from power-law structure in the distribution tail.

To demonstrate the validity of our mean-field approximation, we compare the node degree evolution obtained from a 4^{th} order Runge-Kutta integration of Eq. (3) using our approximation for $F(k, t)$ (see below), against the averaged result from 10^3 realizations of the primary model. As the result, an approximate power law scaling clearly emerges at early evolution stage, and an upper bound to the envelope of node degrees emerges for longer evolution time t necessary to attain larger network sizes (cf. Fig. 6, where the results of the semi-analytical description are based on exponents and prefactors from an approximation of the results of Fig. 7a via Eq. (4)). $F(k, t)$ has a very regular behavior in both variables (k, t) (Fig. 7a) and is accompanied by a node degree distribution $P(k)$ as found for our primary model (Fig. 7b). Over a large range, we can approximate $F(k, t)$ by a power law for small k, and by a second power law at large k:

$$F(k, t) \approx \begin{cases} t^{\alpha} k^{\beta} & \text{if } k \leq k_c \\ k^{\gamma} \dfrac{t^{\alpha} {k_c}^{\beta}}{{k_c}^{\gamma}} & \text{if } k > k_c \end{cases}, \tag{4}$$

where $k_c \sim t^{\lambda}$, and the fractional term for $k > k_c$ simply makes $F(k, t)$ continuous at k_c. The exponents α, β, γ, λ will vary according to the choice of algorithm parameter p, where $0 < \lambda < 1$: i.e. $1 < k_c < t$. In accordance with Fig. 7a, the following observations can be made: First, $\gamma < \beta$ (the exponent of the power law fit decreases as k crosses k_c). Second, $F(t-1, t) = 1$, since $t-1$ is the maximum possible node degree at time t (achieved in Fig. 7a for $t = 25$ only). Similarly, as $p \to 0$, $F(k, t) \to 1$ (the network will tend

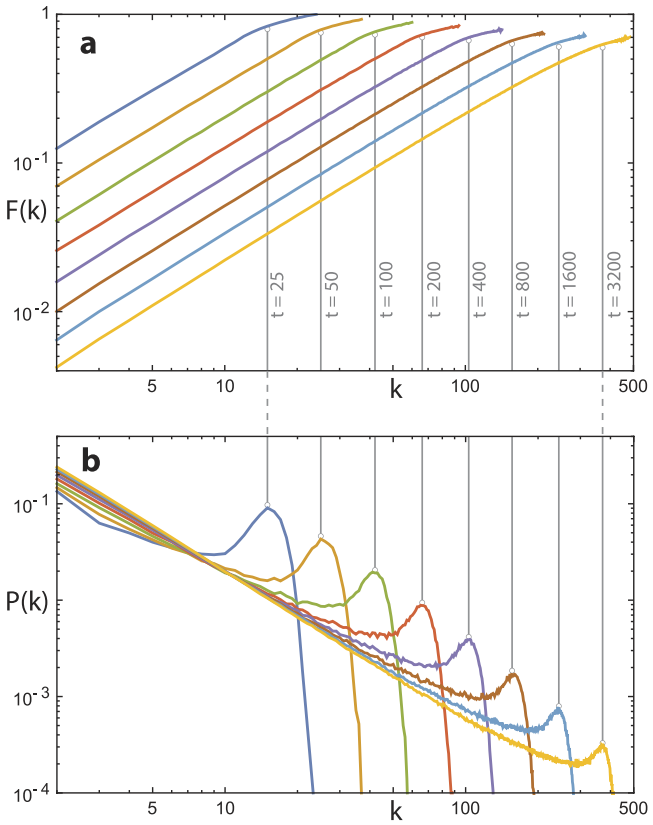


Figure 7. Relation between power-law deviation hump and saturation function: (**a**) Mean field saturation $F(k, t)$, (**b**) mean degree distribution. Data set: 10^3 network realizations for given time t using $p = \frac{1}{24}$. Vertical grey lines are visual aids. The figure indicates the disappearance of the hump structure in the thermodynamic limit.

toward a clique, where all possible connections already exist). When $p = 1$, $F(k, t)$ ceases to be relevant. Finally, for any $p \in (0,1)$, as $t \to \infty$, $F(k, t) \to 0$, since the number of inside edges added at each time-step approximates a constant value, so the network becomes increasingly sparse.

We can use $F(k, t)$ to infer the generated unnormalized degree probability distribution, $N(k, t)$ as follows. Starting from the continuity equation, we may write

$$\frac{\partial}{\partial t}N(k, t) = -\frac{\partial}{\partial k}\left(N(k, t)\frac{\partial k}{\partial t}\right) + \delta_{m,k}, \tag{5}$$

where $\frac{\partial k}{\partial t}$ is given by Eq. (3), and the Kronecker delta function has been included to account for the addition of outside nodes. By differentiating Eq. (3), we notice that Eq. (5) contains the product of k and the derivative of the saturation function F:

$$\frac{\partial}{\partial k}\frac{\partial k}{\partial t} = a_0 + a_1 - a_1\left(k\frac{\partial}{\partial k}F(k, t) + F(k, t)\right), \tag{6}$$

where $a_0 := \frac{m}{2K(t)}$, $a_1 := \frac{(1-p)}{pK(t)}$. The form of $F(k, t)$ implies that a sharp change should occur in the solutions of Eq. (6) around k_c. Indeed, a comparison between $P(k, t)$ and $F(k, t)$ (Fig. 7) supports this suggestion. Thus, we hold the properties of the saturation function $F(k, t)$ responsible for the form of the deviation of $P(k, t)$ from the ideal power law.

Discussion

Examples of edge saturation network growth emerge from the fundamental situation where the state of a physical system is described by a symbol, and where time acting on the states leads to a description in terms of a language (symbolic dynamics and formal languages[20–27], natural languages). Starting with a finite number of N_0 states, observations of the system in time yield sequences of states, that define links on a graph between nodes (states), which implies that more important or more versatile nodes

will have more links. During the refinement of this description, two processes may occur: 1) adjacencies are established between previously unconnected nodes (preferentially between more versatile ones); 2) a new node is added and connected preferentially to already highly connected nodes. Evidently, in many networks there will, however, be a limitation on the number of edges that can be hosted by a given node.

The *Drosophila* courtship body language of 37 fundamental behavioural states[17,18] and its network is an example of such a process. The states are fundamental in the sense that each act could, from the view of the physics of body motion, be followed by any other act. Some transitions, however, are generally not taken, leading to edges missing. Well-defined connected sub-networks characterize a chosen courtship partner's class, according to which protagonists can be distinguished (male, female (virgin, mature, mated), fruitless). Within these bounds, courtship exploits the available expression space, corroborating the view that it might advertise individual properties of the sender into the eyes of a courtship partner[18,28]. To compare our network growth algorithm with the data from male-female interaction, we grow the network until the number of nodes (symbols) is depleted, with p chosen so that on average the number of edges matches that of the courtship network. A comparison -without further fitting- exhibits that the two degree distributions match extremely well and that the proposed generating algorithm is very specific (Fig. 5).

Our paradigm may also appear in the guise of an equilibrium condition in the following sense. Complex networks in physics or in biology are often constrained to maintain some 'average' conditions. As soon as (possibly: self-enhancing) node interaction sets in, this needs to be balanced by homeostasis, i.e. a competitive, counter-balancing mechanism that weakens other connections of the same node to the network[8]. In the neural networks domain, a closely related principle is known as 'Hebbian learning'[29]. Self-organized Hebbian-learning[30] in the super-paramagnetic[31] phase of ensembles has been proven a reliable and efficient way of clustering that does away with convexity requirements of cluster borders[32]. A very similar approach has also been used as a synchronization model for coupled oscillators, where the oscillators' struggle to synchronize is expressed by competing connection strengths w_{ij} that evolve according to the dynamical update rule $\frac{dw_{ij}}{dt} = s_{ij} - w_{ij}\left(\sum_{(i,k)\in E} s_{ik}\right)$ [8], where s_{ij} measures the pairwise oscillator synchrony. The resulting distribution of w_{ij} has been shown to tend for intermediate coupling strengths towards a hump-terminated power-law (cf. Fig. 2a). This dynamical law expresses the limited resources available for the local wiring around each node, which in our model is encoded in the probability p ruling the edge saturation. We envisage that also avalanche distributions of the typical form of Fig. 2a could be understood similarly[11].

Many interesting real-world phenomena dwell on the mesoscale. In social networks, the largest scale is relevant, e.g., for the study of disease and rumour spreading, but more subtle social dynamics happens within the community structures[33,34]. Our results suggest that a large class of systems can be formulated as growing along simple principles, similar and in addition to preferential attachment. The sets of m, p parameters needed to recover an experimental distribution, i.e. the violation of the ideal power law on the macroscopic scale, provides us with an insight about the local mesoscale structures present in the network. In this way, starting from non-ideal power law distributions of complex networks, an avenue opens towards the identification and understanding of interesting mesoscale real-world phenomena in physics.

References

1. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. Complex networks: structure and dynamics. *Phys. Rep*. **424,** 175 (2006).
2. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74,** 47 (2002).
3. Cohen, R. & Havlin, S. Complex networks: structure, robustness and function. (Cambridge University Press, Cambridge, England, 2010).
4. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286,** 509 (1999).
5. Amaral, L. A. N., Scala, A., Barthélémy, M. & Stanley, H. E. Classes of small-world networks. *Proc. Natl. Acad. Sci. USA*. **97,** 11149 (2000).
6. Mossa, S., Barthélémy, M., Stanley, H. E. & Amaral, L. A. N. Truncation of power law behavior in "scale-free" network models due to information filtering. *Phys. Rev. Lett.* **88,** 138701 (2002).
7. Dorogovtsev, S. N. & Mendes, J. F. F. Language as an evolving word web. *Proc. R. Soc. Lond. B* **268,** 2603 (2001).
8. Assenza, S., Gutiérrez, R., Gómez-Gardeñes, J., Latora, V. & Boccaletti, S. Emergence of structural patterns out of synchronization in networks with competitive interactions. *Sci. Rep*. **1,** 99 (2011).
9. Eurich, C. W., Herrmann, J. M. & Ernst, U. A. Finite-size effects of avalanche dynamics. *Phys. Rev. E* **66,** 066137 (2002).
10. Levina, A., Herrmann, J. M. & Geisel, T. Dynamical synapses causing self-organized criticality in neural networks. *Nat. Phys*. **3,** 857 (2007).
11. de Arcangelis, L., Lombardi, F. & Herrmann, H. J. Criticality in the brain. *J. Stat. Mech*. **3,** P03026 (2014).
12. Del Genio, C. I., Gross, T. & Bassler, K. E. All scale-free networks are sparse, *Phys. Rev. Lett.* **107,** 178701 (2011).
13. Dorogovtsev, S. N., Mendes, J. F. F. & Samukhin, A. N. Size-dependent degree distribution of a scale-free growing network. *Phys. Rev. E* **63,** 062101 (2001).
14. Guimaraes, P. R., de Aguiar, M. A. M., Bascompte, J., Jordano, P. & dos Reis S. F. Random initial condition in small Barabasi-Albert networks and deviations from the scale-free behavior. *Phys. Rev. E* **71,** 037101 (2005).
15. Waclaw, B. & Sokolov, I. M. Finite-size effects in Barabási-Albert growing networks. *Phys. Rev. E* **75,** 056114 (2007).
16. Schneider, C. M., de Arcangelis, L. & Herrmann, H. Scale-free networks by preferential depletion. *Euro. Phys. Lett*. **95,** 16005 (2011)
17. Stoop, R. & Arthur, B. Periodic orbit analysis demonstrates genetic constraints, variability, and switching in Drosophila courtship behavior. *Chaos* **18,** 023123 (2008).

18. Stoop, R. & Joller, J. Mesocopic Comparison of Complex Networks Based on Periodic Orbits. *Chaos* **21,** 016112 (2011).
19. Dorogovtsev, S. N. & Mendes, J. F. F. Evolution of Networks. (Oxford University Press, Oxford, 2003).
20. Grassberger, P. & Kantz, H. Generating partitions for the dissipative Hénon map. *Phys. Lett. A* **113,** 235 (1985).
21. Cvitanović, P., Gunaratne, G. H. & Procaccia, I. Topological and metric properties of Hénon-type strange attractors. *Phys. Rev. A* **38,** 1503 (1988).
22. Bai-Lin, H. Elementary Symbolic Dynamics and Chaos in Dissipative Systems (World Scientific, Singapore, 1989).
23. Stoop, R. Bivariate thermodynamic formalism and anomalous diffusion. *Phys. Rev. E* **49,** 4913 (1994).
24. Stoop, R. & Parisi, J. Evaluation of probabilistic and dynamical invariants from finite symbolic substrings-comparison between two approaches. *Physica D* **58,** 325 (1992).
25. Stoop, R. Phase transitions in the approximated and asymptotic generalized entropy spectrum of a nonhyperbolic system. *Phys. Rev. A* **46,** 7450 (1992).
26. Lai, Y.-C., Bollt, E. & Grebogi, C. Communicating with chaos using two-dimensional symbolic dynamics. *Phys. Lett. A* **255,** 75 (1999).
27. Klages, R. Microscopic chaos, fractals and transport in non-equilibrium statistical mechanics (World Scientific, Singapore, 2007).
28. Stoop, R., Nüesch, P., Stoop, R. L. & Bunimovich, L. A. At grammatical faculty of language, flies outsmart men. *PLoS ONE* **8,** e70284 (2013).
29. Hebb, D. The Organization of Behavior (Wiley & Sons, New York, 1949).
30. Landis, F., Ott, T. & Stoop, R. Hebbian self-organizing integrate-and-fire networks for data clustering. *Neur. Comp.* **22,** 273 (2010).
31. Ott, T. *et al.* Sequential superparamagnetic clustering for unbiased classification of high-dimensional chemical data. *J. Chem. Inf. Comput. Sci.* **44,** 1358 (2004).
32. Gomez, F., Stoop, R. L. & Stoop, R. Universal dynamical properties preclude standard clustering in a large class of biochemical data. *Bioinformatics* **30,** 2486 (2014).
33. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*. **99,** 7821 (2002).
34. Palla, G., Derényi, I., Farkas, I. & Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435,** 814 (2005).
35. Yanai, I., Camacho, C. J. & DeLisi, C. Predictions of gene family distributions in microbial genomes: evolution by gene duplication and modification. *Phys. Rev. Lett.* **85,** 2641 (2000).

Acknowledgements

This work was supported by the Swiss National Science Foundation (Grant 200021-153542/1 to R.S.).

Author Contributions

R.S. and T.L. designed the research, T.L. and F.G. carried out the analysis, R.S. wrote the manuscript.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Lorimer, T. *et al.* Two universal physical principles shape the power-law statistics of real-world networks. *Sci. Rep.* **5**, 12353; doi: 10.1038/srep12353 (2015).